Sociodynamic Discrete Choice on Networks in Space: Role of Utility Parameters and Connectivity in Emergent Outcomes

Elenna R. Dugundji a,*, László Gulyás b

a Universiteit van Amsterdam, P.O. Box 16697, 1001 RD Amsterdam, Netherlands
b AITIA International Inc., Czet János u. 48-50, 1039 Budapest, Hungary

Abstract

We consider a model where a commuter’s transportation mode choice is directly influenced by percentages of neighbors and socioeconomic peers making each choice. Discrete-choice estimation results controlling overall mechanisms related to individual heterogeneous preferences are embedded in a multi-agent based simulation in order to observe the evolution of choice behavior over time with sociodynamic feedback. We find that the estimated utility parameters for different plausible sociogeographic network scenarios can generate dramatically different dynamics. However, in a hypothetical experiment we find that swapping sociogeographic networks does not significantly change the long-run outcome of the simulation, when utility parameters are held fixed.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [name organizer]

Keywords: Multi-agent based social simulation, Discrete choice, Social networks, Spatial interaction, Travel demand

1. Introduction

There is growing awareness and interest in the influence that social factors have on transportation and land use behaviors [1,2]. We consider a model where a commuter’s transportation mode choice is influenced by percentages of neighbors and socioeconomic peers making each choice. Such inter-household feedback can have very important implications for the prediction of (system-wide) results over the course of time. If such feedback exists, it can namely propel or hinder the adoption of a mode over time [3]. In diverse literature this dynamically reinforcing behavior is referred to as a social multiplier, a cascade, a bandwagon effect, imitation, contagion, herd behavior, etc. [4]. Our work extends theoretical

* Corresponding author. Tel./fax: +31 20 692 5813.
E-mail address: e.r.dugundji@uva.nl
work on multinomial choice with social interactions [5] in three ways. First, we allow for the possibility of common unobserved attributes of the choice alternatives using the nested logit model [6]. Second, a key feature of our work is that we explicitly consider non-global interactions, with several different social and spatial network structures. Third, additional heterogeneity is introduced in the model through different mechanisms, such as individual-specific sociogeographic characteristics of the commuters as well as individual-specific attributes of the choice alternatives, and the availability of alternatives.

We present an exploratory application of the model to transportation mode choice using pseudo-panel microdata collected in the greater Amsterdam region. Here we combine econometric estimation with computational techniques from the field of multi-agent based simulation. This paper extends previous work by the authors [7] by further studying the role of the utility parameters and connectivity of sociogeographic networks on the emergent outcomes of the multi-agent based simulation. We also highlight limitations of our present study in any extension for policy considerations on the adoption of innovation in transportation mode choice. Finally, we suggest recommendations for future work.

2. Sociogeographic Networks

The research reported here explores interactions between a decision-maker and the aggregate actions of other decision-makers proximally situated in a spatial network, and interactions between a decision-maker and the aggregate actions of other decision-makers associated in a socioeconomic network. We use a priori beliefs about the social and/or spatial dimension of interactions to formulate the connectivity of the network. Technically, however, interactions between identifiable decision-makers may also be modeled using the approach described in this paper given the availability of suitable data, and thus methods reported here may prove to be useful in those areas as well.

In the case study to be discussed, we have rich socioeconomic data for each respondent as well as the geographic location of each respondent’s residence and work location. This allows us to define aggregate interactions by grouping agents into geographic neighborhoods or into socioeconomic groups where the influence is assumed to be more likely. In the simplest case, these groups are assumed to be mutually exclusive and collectively exhaustive. That is each agent belongs to one and only one group. The agent is assumed to be influenced by the average choice behavior of his or her group, and the influence by other groups is assumed to be negligible. At a global level, the picture is a fragmented or disconnected network of clustered groups. If we are interested in equilibrium behavior, the consequences of such an assumption are important: there is no transmission of influence across groups, and the global picture is a weighted average behavior of the separate clusters. Thus we also consider cases with overlapping groups, with agents for example connected by social group as well as by residential district, or by postcode regions of residence and work location. This leads to a giant cluster for the empirical examples under consideration, with the important implication that influence can spread throughout the entire population.

3. Case Study

The data used in this paper originates from travel questionnaires administered by the Municipality of Amsterdam Agency for Infrastructure, Traffic and Transport, in Amsterdam and a neighboring suburb to the south of the city, Amstelveen. The data set made available by the Agency is a subset of the full modal split database, containing direct home-work trips and direct work-home trips where the purpose of the trip at the non-home location is classified as either “work” or “business.” Geographical location is given in terms of the centroid of a traffic analysis zone. The data received includes records of trips where respondents have indicated one of the following transportation mode choices: external system public transit or internal system public transit (23,7% mode share); bicycle or moped/motorcycle (26,7% mode share); car driver or car passenger (49,6% mode share). The final sample used in the case study contains
2913 respondents. Raw variables available for use in the model are availability of public transit, car ownership, gender, income category, education level, age, in-vehicle and out-of-vehicle travel time for public transit, travel time by bicycle, and travel time and parking time for car. Availability of bicycle is generated based on a 75 minute travel time cut-off.

3.1. Definition of Interaction Variables

Now we turn to the specification of the network interdependence. We begin with a broad classification by residential district applying the 9 districts represented in the sample. Next using the three variables age, income and education, 13 socioeconomic groups are defined [3,7]. Finally, to be able to test the effect of spatial scale, we define a smaller neighborhood region of influence on the basis of 4-digit postcode using the 67 postcode regions represented in the sample. We may hypothesize that the smaller spatial scale network interdependence defined by postcode may be more homogeneous with regard to choice behavior than that for the variables defined on the basis of district. Thus we may expect the coefficient on these variables to be relatively stronger. Network interaction variables for four scenarios are defined. Two scenarios consider clustered groups: Residential district; Social group. Two scenarios consider overlapping groups: District and Social group; Postcode and Social group. For purposes of comparison across the four scenarios, in the treatments where social and spatial interdependence are considered jointly, agents are assumed not to distinguish between their socioeconomic peers’ and their fellow district residents or neighbors when considering their choice behavior.

3.2. Specification of Utility Functions

A trinary transportation mode choice model to work is estimated using the freely available, open source optimization toolkit Biogeme [8]. Various piecewise linear specifications of all travel time related variables as well as age were tested against linear, quadratic and logarithmic forms of these variables. Considering various a priori hypotheses of behavior in the region and after statistical comparison of the alternative nonlinear specifications of variables against the linear versions thereof using loglikelihood ratio tests and non-nested tests [9], a baseline multinomial logit model is estimated. Estimation of three successive nested logit models first with public transit nested with bicycle, then with public transit nested with car, and finally with bicycle nested with car, show the first nesting structure to be most significant in terms of loglikelihood ratio test and in terms of the a t-test on the nest coefficient. The third nesting structure was not indicated. The nested logit model thus adds one additional parameter to the multinomial specification, namely the scale parameter for the transit-bicycle nest. Finally we consider the network interaction variables for the four scenarios described in section 3.1.

We conclude from t-tests on the network interaction variables, that for this particular case study and the network definitions under consideration, systematic field effects representing social and spatial network interactions between an agent and the aggregate behavior of other reference agents do indeed have explanatory power [10]. On the basis of non-nested model specification tests, we find the fit for overlapping postcode and social group is best, as expected. The fit for broad district clusters alone is worst. Interestingly, there is no statistically significant gain in fit at the 0.05 level between the scenario with social group clusters versus the scenario with overlapping district and social group. In light of the latter finding, we will find that the emergent outcomes over time when these models are embedded in a multi-agent based simulation with feedback are particularly noteworthy.

For continuity in the model development process extending the original discrete choice with interactions research by Brock and Durlauf [5], a nested logit model is considered in this paper. However, an important econometric issue arises in the empirical estimation of discrete choice models using a nested logit specification in that, while unobserved heterogeneity is accounted for across alternatives, the Gumbel
error terms are still assumed to be identically and independently distributed across decision-makers. It is not obvious that this is a valid assumption when we are specifically considering interdependence between decision-makers’ choices [3]. For the purposes of this paper, we accept that the estimated values may be biased. We are interested here in getting an idea methodologically under what conditions a runaway effect is generated and what influences this. It is very useful then to understand the dynamic behavior of a simple nested logit model, before proceeding to understand the dynamic behavior of models with even more complex kernels. Such an understanding built-up step-by-step is important both theoretically and conceptually as well as for good practice in multi-agent based simulation [11].

4. Results and Discussion

Using the RePast agent-based modeling platform (http://repast.sourceforge.net), we create a computational version of our nested logit models with heterogeneous agents and sociogeographic network interaction. Discrete-choice estimation results controlling overall mechanisms related to individual heterogeneous preferences are embedded in the multi-agent based model to be able to observe the simulated evolution of choice behavior over time with sociodynamic feedback due to network effects. Example results for different random seeds are shown in Fig. 1. Each run is allowed to iterate for 600,000 time steps, or roughly about 200 revisions of choices with asynchronous decision-making for the sample size of 2913 agents.

Fig.1. Observed long-run mode shares for multi-agent simulation of nested logit models with social feedback on different sociogeographic networks

There are several immediately striking features of the long-run results. First, we notice that in all scenarios, the long-run mode shares in Fig. 1 moved significantly away from the initial overall modal split (23.7% public transit share; 26.7% bicycle or moped/motorcycle share; 49.6% auto driver or auto passenger share). Second, we notice that the long-run results are also fairly stable: there is little variation in the long-run results for a given scenario. This is true for both scenarios with overlapping groups where influence has the possibility, in principle, to spread through the entire sample. Since it is also true for both scenarios with disconnected clusters where there is no possibility for transmission of influence across groups, this implies that the modal split within the clusters was effectively the same across clusters for a given scenario. Third, we notice in all cases the auto share strongly decreased. This is especially
remarkable since the auto mode had initially a share about twice as large as either of the other modes. We see that the feedback effect was thus indeed significant in dynamically hindering the auto mode in the long-run in all scenarios in a well-defined manner.

What is curious is that the feedback effect one hand dynamically propels the transit mode for the case of network interaction by residential district clusters, and by overlapping residential district and social group, and on the other hand dynamically propels the bicycle mode for the case of network interaction by social group clusters, and by overlapping postcode and social group. This is a dramatic difference, emphasizing how important it would be in an application for policy purposes to know in the case of clusters whether influence actually works through neighbors or through socio-economic peers, or in the case of overlapping groups what the regional scale is of neighborhood influence. We can in any case conclude resoundingly: if a feedback effect can be assumed, the precise details of the connectivity sociographic networks matter!

It is important to recognize however that there are two stages in our process where the socio-graphic network enters. First, the network enters in the econometric estimation in determining the value of the estimated coefficients. Second, the network enters in the multi-agent based simulation in determining the course of the spread of influence when the feedback is strong enough. We may wonder then what is the driving factor of the results: is it simply the strength of the feedback effect relative to the other components of the utility? or is it the connectivity of the network during the transmission process? or both? For example, if a feedback effect can be assumed, in a campaign to promote a particular mode or new service, we would want to know whether to focus efforts on the way the mode is promoted to make the adoption most convincing, or whether to focus for example, on seeding opinion-makers to try to influence the connectivity of the socio-graphic network.

To gain some insight to the answer with regard to this particular case study, we run a hypothetical simulation experiment with sociogeographic networks swapped, while holding the utility parameters fixed [10]. We find that only in the case of the social group parameters did the connectivity of the network seem to have some slight effect on the outcome of the multi-agent simulation. In our particular case study, we conclude that the strength of the feedback effect relative to the other components of the utility is the dominant factor in generating the long-run results. That is, in our particular case study, the connectivity appears not to be very relevant at the transmission stage. This said, it is important to note that the networks studied here are fairly dense by definition, due to the nature of the aggregate interaction assumed within groups. Earlier work by the authors [11] on a simple binary choice model with social interactions on abstract classes of networks over a sweep of network density, indicated that sparse networks were more sensitive in the outcomes of transmission.

5. Conclusions

We have extended previous work on discrete choice with social interactions in important ways. We consider a model where an agent’s choice is directly influenced by the percentages of the agent’s neighbors and socio-economic peers making each choice, under four different scenarios. Two scenarios depict influence within a disconnected network of clustered groups. Two scenarios depict influence within overlapping social and spatial groups. Given the availability of appropriate data, our approach is principle directly extendable to the identifiable agent case. We observe that the estimated utility parameters for different hypothetical sociogeographic network scenarios can generate dramatically different dynamics. This finding underscores the need for more empirical research to understand actual sociogeographic influence networks [12-17].

A challenging direction of on-going work by the authors addresses evolving networks in coupling with the evolving behavioral dynamics [18]. A motivation for this direction of work is to be able to account for residential mobility, occupational mobility and other life cycle changes in social-spatial networks.
impacting transportation mode choice [19]. The econometric aspects of estimating utility parameters for a residential location choice model with social and spatial network interactions are non-trivial and data intensive [20].

Acknowledgements

The authors would like to gratefully acknowledge discussion with Harry Timmermans, Theo Arentze, Cars Hommes, Frank le Clercq, Loek Kapoen, George Kampis, József Váncza and András Márkus. Special thanks are also due to Guus Brohm and Nelly Kalfs at the Agency for Infrastructure, Traffic and Transport of the Municipality of Amsterdam, and to Willem Vermin and the High Performance Computing support team at SARA Computing and Networking Services, Science Park Amsterdam.

References